SpiderFoot OSINT Tool - Complete Project Guide

Table of Contents

1. Introduction
2. What is SpiderFoot?
3. Key Features
4. System Requirements
5. Installation Guide
6. Configuration
7. Usage Guide
8. Command Reference
9. Practical Use Cases
10. Advantages and Disadvantages
11. Best Practices
12. Legal and Ethical Considerations
13. Resources

Introduction

This comprehensive guide provides everything you need to understand, install, configure, and effectively use SpiderFoot for Open Source Intelligence (OSINT) gathering. Whether you're a security professional, penetration tester, researcher, or cybersecurity enthusiast, this guide will walk you through every aspect of SpiderFoot.

What is SpiderFoot?

SpiderFoot is an open-source intelligence (OSINT) automation tool that integrates with multiple data sources to gather information about a target. It automates the process of collecting intelligence from over 200 public data sources including WHOIS records, DNS records, search engines, social media platforms, threat intelligence feeds, and more.

Definition

SpiderFoot is a reconnaissance tool designed to automate the OSINT gathering process by:

- Querying multiple data sources simultaneously
- Correlating discovered data to find relationships
- Presenting findings in an organized, actionable format
- Identifying potential security risks and exposure points

Purpose and Importance

Why SpiderFoot Matters:

1. **Comprehensive Reconnaissance**: Provides a 360-degree view of your digital footprint or target
2. **Time Efficiency**: Automates hours of manual OSINT work into minutes
3. **Threat Intelligence**: Identifies potential vulnerabilities and security exposures
4. **Attack Surface Mapping**: Discovers all internet-facing assets associated with a target
5. **Data Correlation**: Connects disparate pieces of information to reveal hidden relationships

System Requirements

Minimum Requirements

- **Operating System**: Linux, Windows, macOS
- **Python Version**: Python 3.7 or higher
- **RAM**: 2GB minimum (4GB recommended)
- **Disk Space**: 500MB for installation
- **Network**: Stable internet connection

Recommended Environment

- **OS**: Kali Linux, Ubuntu 20.04+, or Debian-based distributions
- **RAM**: 8GB
- **Processor**: Multi-core CPU
- **Network**: High-speed broadband connection

Installation Guide

Method 1: Installation on Linux (Ubuntu/Debian)

Step 1pi: Update System Packages

bash

sudo apt update && sudo apt upgrade -y

Step 2: Install Required Dependencies

bash

sudo apt install python3 python3-pip git -y

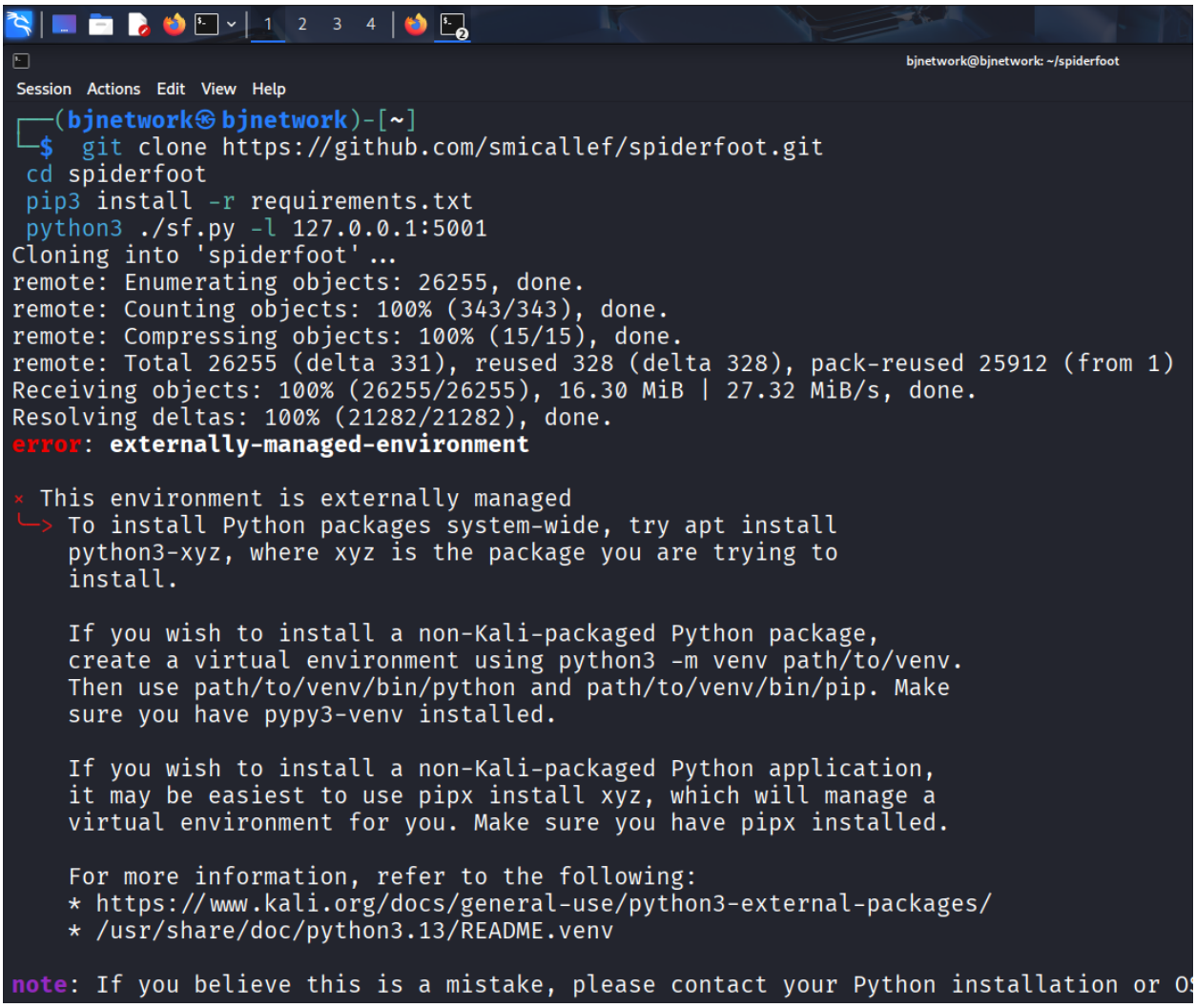
```
┌──(bjnetwork㉿bjnetwork)-[~/spiderfoot]
└─$ sudo apt install python3 python3-pip git -y
python3 is already the newest version (3.13.7-1+b1).
python3-pip is already the newest version (25.3+dfsg-1).
python3-pip set to manually installed.
git is already the newest version (1:2.51.0-1).
git set to manually installed.
```

Step 3: Clone SpiderFoot Repository

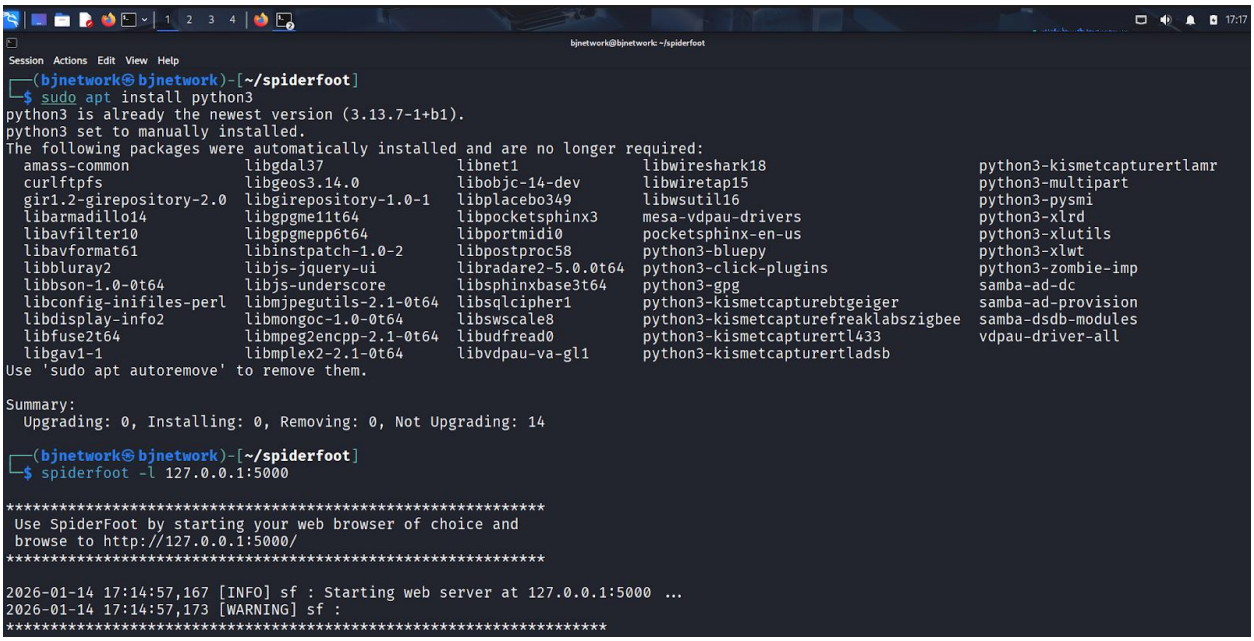
cd ~

git clone https://github.com/smicallef/spiderfoot.git

cd spiderfoot

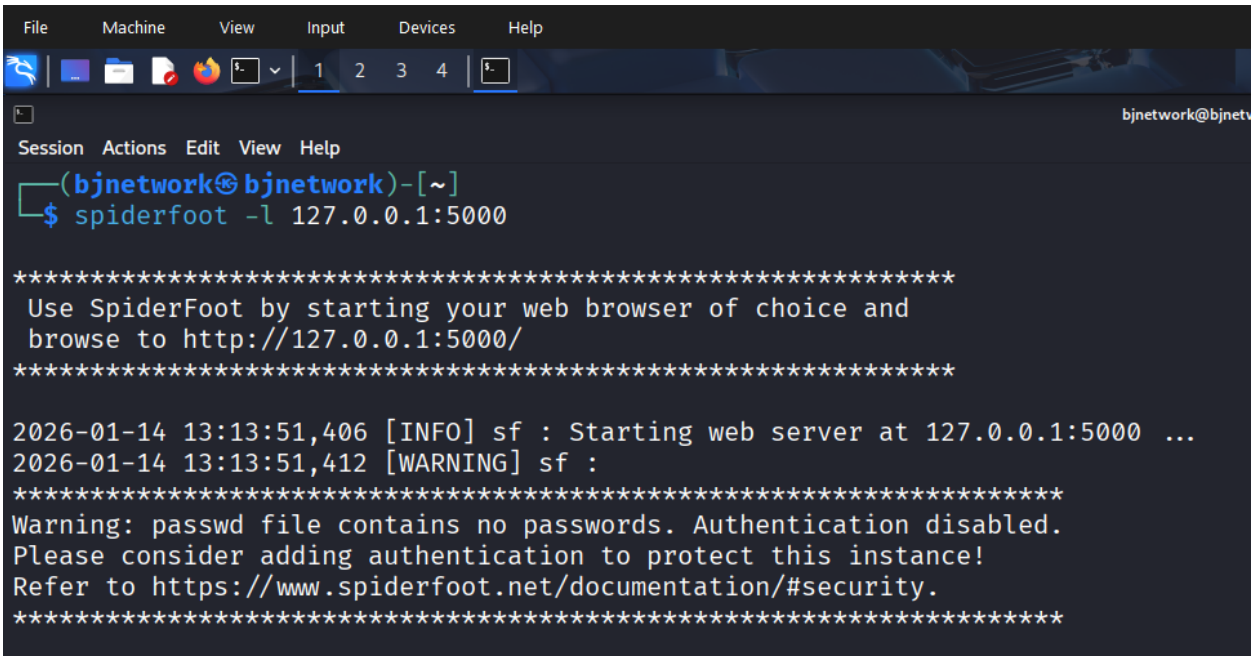


Step 4: Install Python Requirements



Step 1: Start SpiderFoot

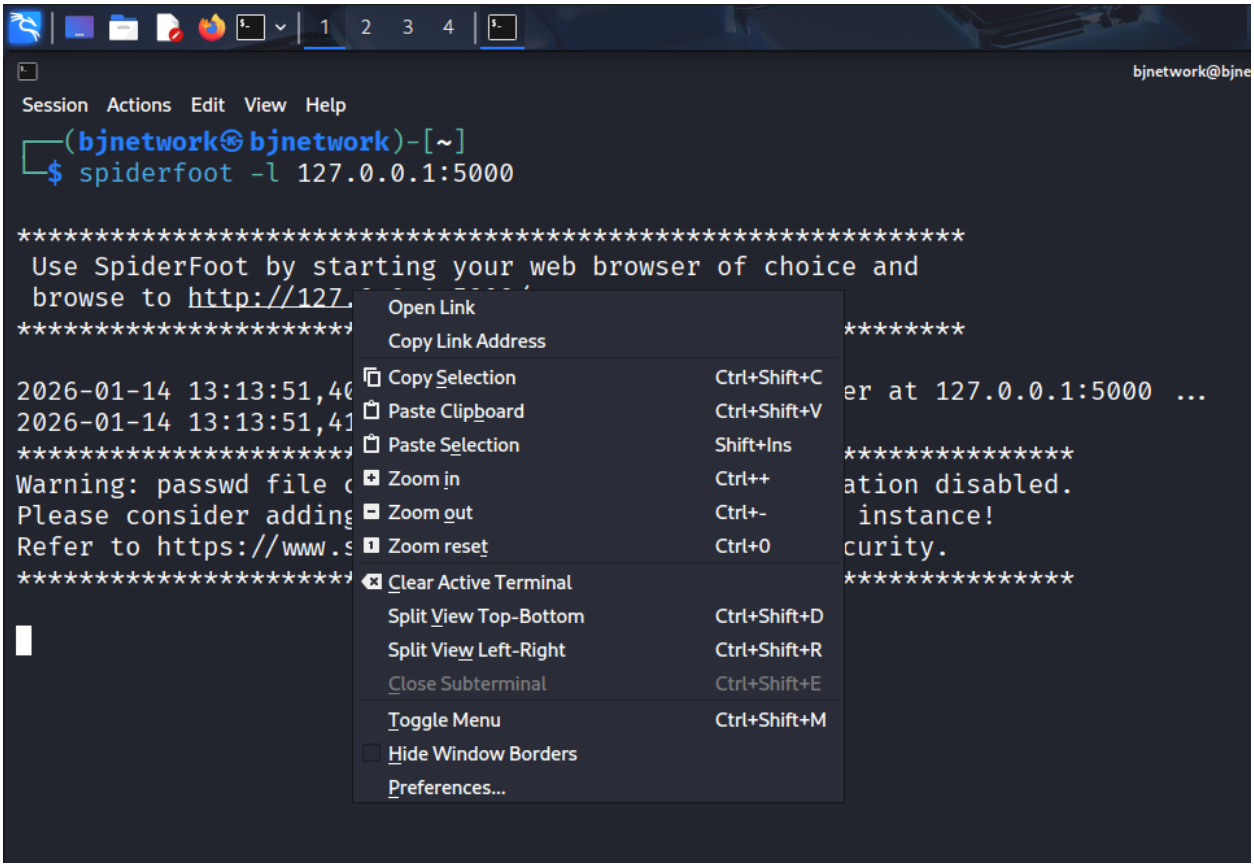
spiderfoot -l 127.0.0.1:5000

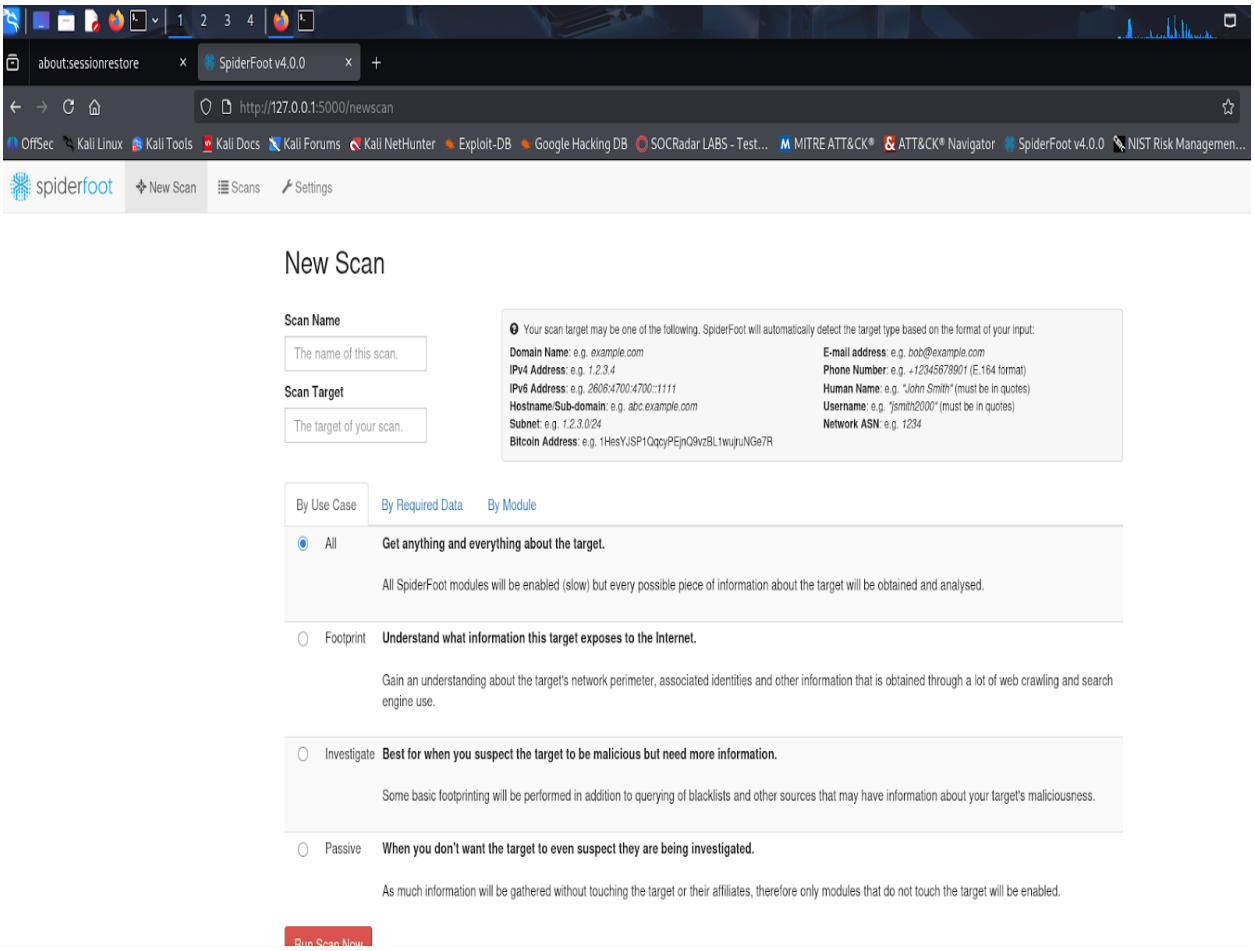


Step 6: Access Web Interface

Open your browser and navigate to:

http://127.0.0.1:5001





Method 2: Docker Installation

Step 1: Install Docker

Follow instructions at docker.com

Step 2: Pull SpiderFoot Docker Image

bash

docker pull spiderfoot/spiderfoot

Step 3: Run SpiderFoot Container

bash

docker run -p 5001:5001 spiderfoot/spiderfoot

Step 4: Access Interface

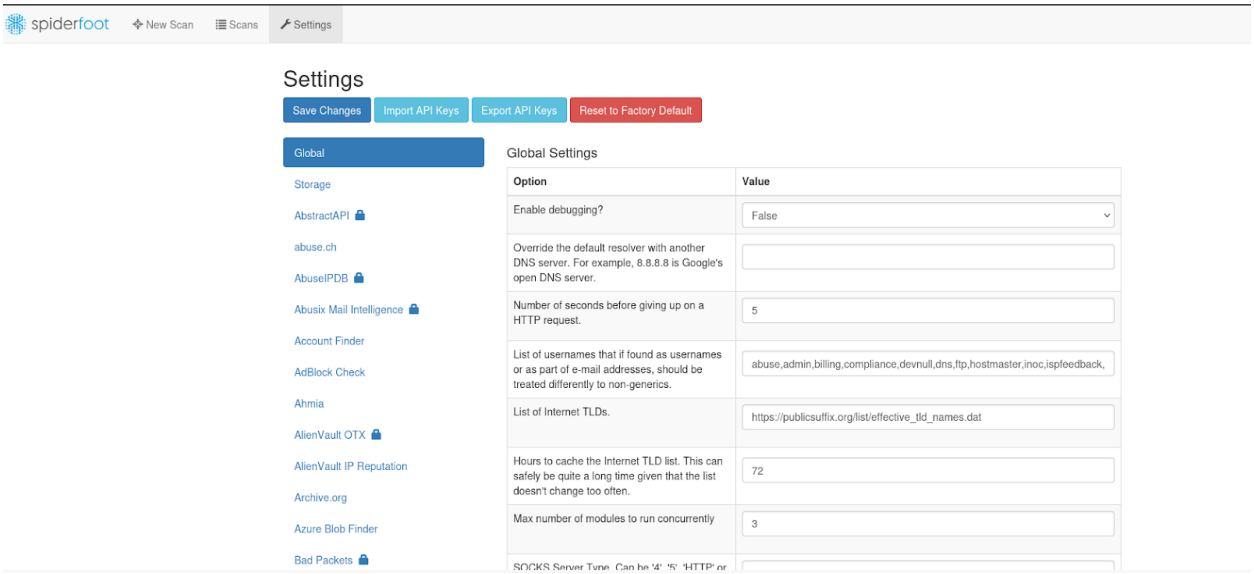
Navigate to http://localhost:5001

Configuration

Initial Setup

Step 1: Access Settings

Once SpiderFoot is running, click on **Settings** in the web interface.



Step 2: Configure API Keys (Optional but Recommended)

SpiderFoot works better with API keys for various services:

- **Shodan API**: For device and service discovery
- **VirusTotal API**: For malware and threat intelligence
- **HaveIBeenPwned API**: For breach data
- **AlienVault OTX API**: For threat intelligence
- **Censys API**: For internet-wide scanning data

How to add API keys:

1. Navigate to Settings → API Keys
2. Enter your API key for each service
3. Click Save

Step 3: Module Configuration

Enable or disable specific modules based on your needs:

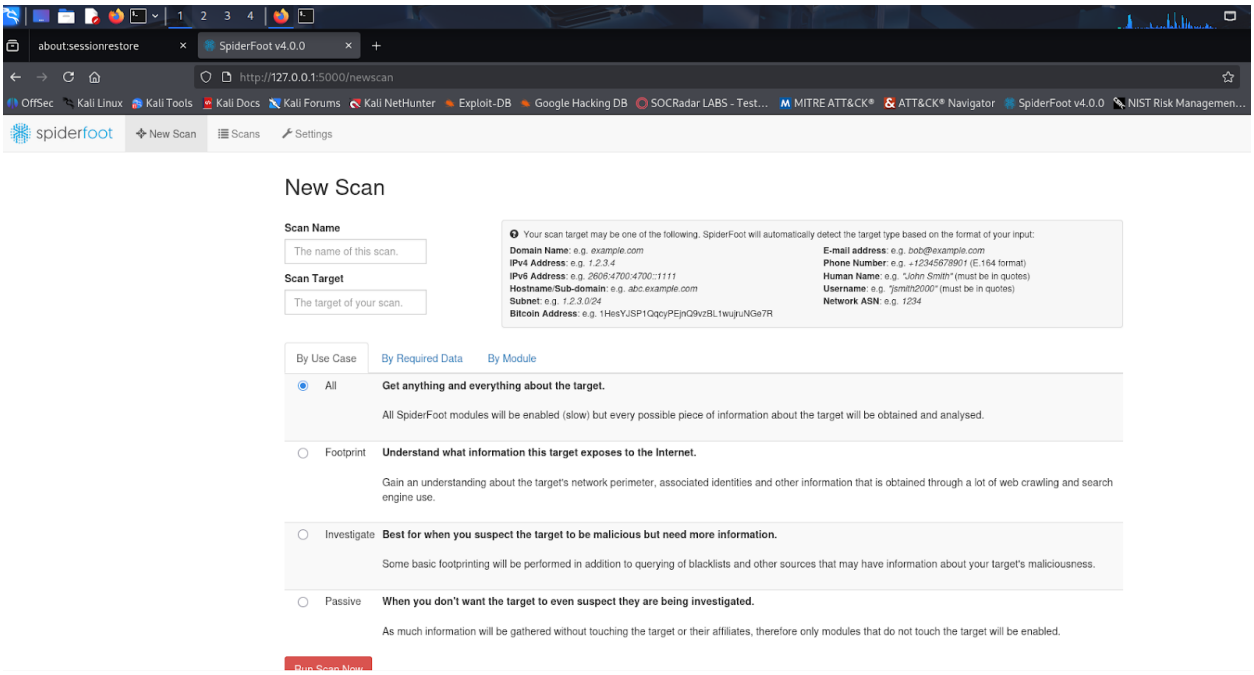
1. Go to Settings → Modules
2. Check/uncheck modules you want to use
3. Save configuration

Usage Guide

Creating Your First Scan

Step 1: Start a New Scan

1. Click on **New Scan** from the main interface
2. Enter scan name (e.g., "Target Domain Reconnaissance")



Step 2: Define Target

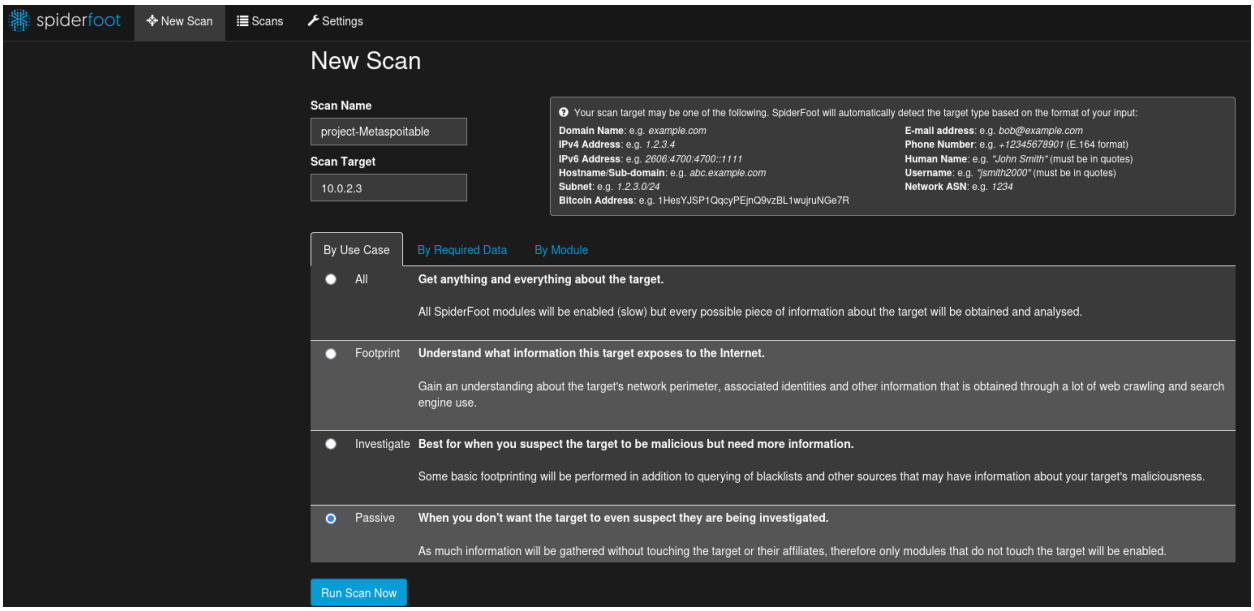
Enter your target in one of these formats:

- **Domain**:
- **IP Address**:
- **Email Address**:
- **Person Name**:
- **Phone Number**:
- **Username**:

Step 3: Select Scan Type

Choose from predefined scan profiles:

- **All**: Uses all available modules (comprehensive but slow)
- **Footprint**: Basic footprint discovery
- **Investigate**: Deep investigation with correlation
- **Passive**: Only passive reconnaissance (no direct contact)



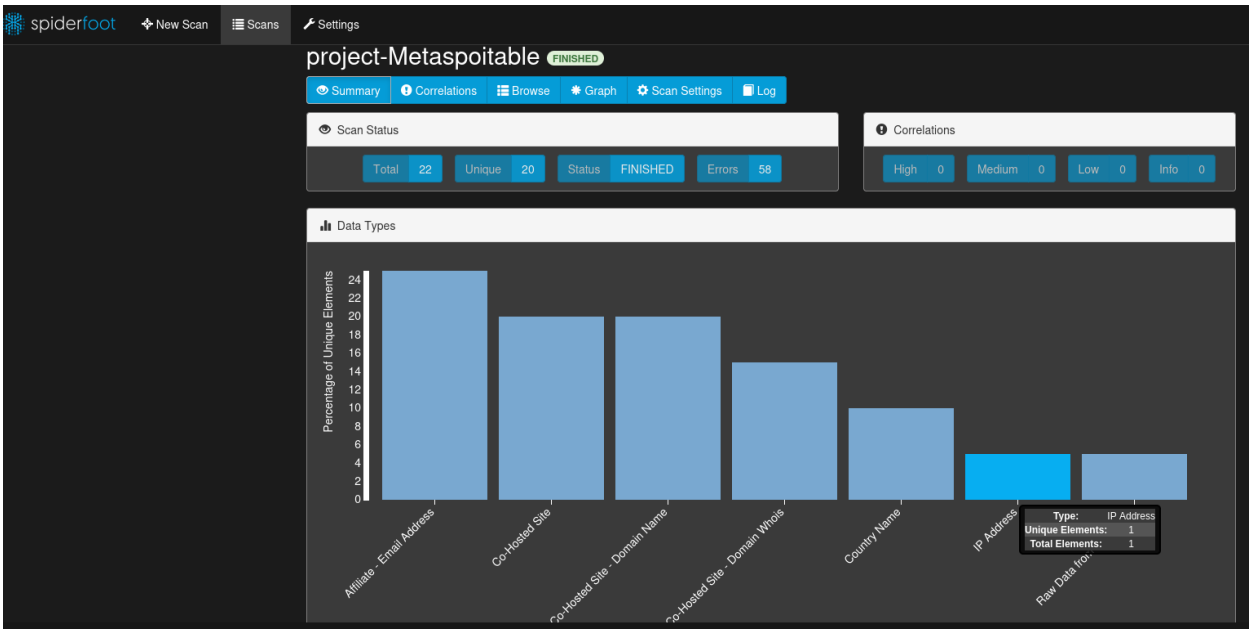
Step 4: Customize Modules (Optional)

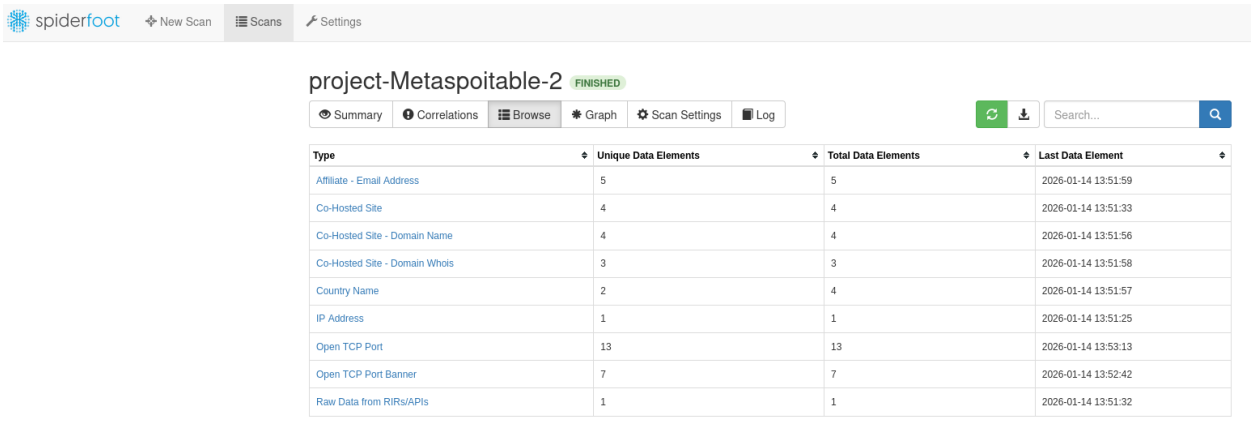
Click **Show Advanced Options** to:

- Select specific modules
- Adjust timeout settings
- Configure thread count

Step 5: Start Scan

Click **Run Scan** button



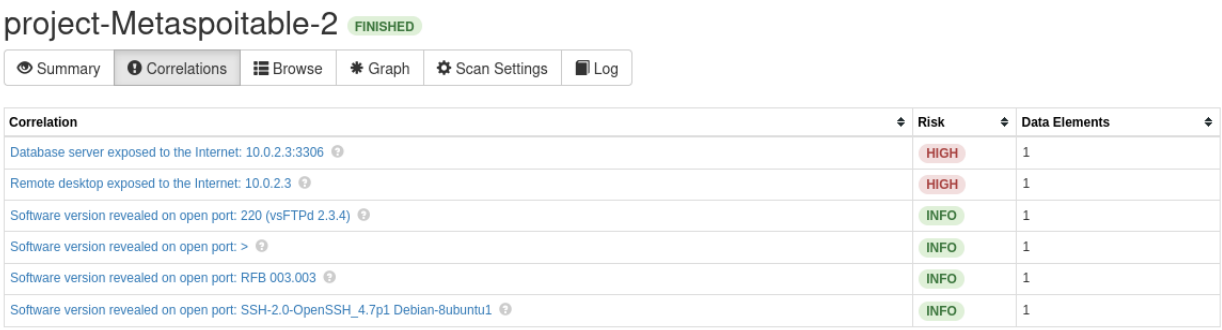


Type	Unique Data Elements	Total Data Elements	Last Data Element
Affiliate - Email Address	5	5	2026-01-14 13:51:59
Co-Hosted Site	4	4	2026-01-14 13:51:33
Co-Hosted Site - Domain Name	4	4	2026-01-14 13:51:56
Co-Hosted Site - Domain Whois	3	3	2026-01-14 13:51:58
Country Name	2	4	2026-01-14 13:51:57
IP Address	1	1	2026-01-14 13:51:25
Open TCP Port	13	13	2026-01-14 13:53:13
Open TCP Port Banner	7	7	2026-01-14 13:52:42
Raw Data from RIRs/APIs	1	1	2026-01-14 13:51:32

project-Metaspoitable-2 FINISHED

Summary | Correlations | Browse | Graph | Scan Settings | Log

Time	Component	Type	Event
2026-01-14 13:53:12	sfp_portscan_tcp	STATUS	TCP port 10.0.2.3:513 found to be OPEN.
2026-01-14 13:53:12	sfp_portscan_tcp	STATUS	TCP port 10.0.2.3:111 found to be OPEN.
2026-01-14 13:52:42	sfp_email	DEBUG	Received event, TCP_PORT_OPEN_BANNER, from sfp_portscan_tcp
2026-01-14 13:52:42	sfp_dnsresolve	DEBUG	Received event, TCP_PORT_OPEN_BANNER, from sfp_portscan_tcp
2026-01-14 13:52:42	sfp__stor_db	DEBUG	Storing an event: TCP_PORT_OPEN_BANNER
2026-01-14 13:52:42	sfp__stor_db	DEBUG	Storing an event: TCP_PORT_OPEN
2026-01-14 13:52:42	sfp__stor_db	DEBUG	Storing an event: TCP_PORT_OPEN



Monitoring Scan Progress

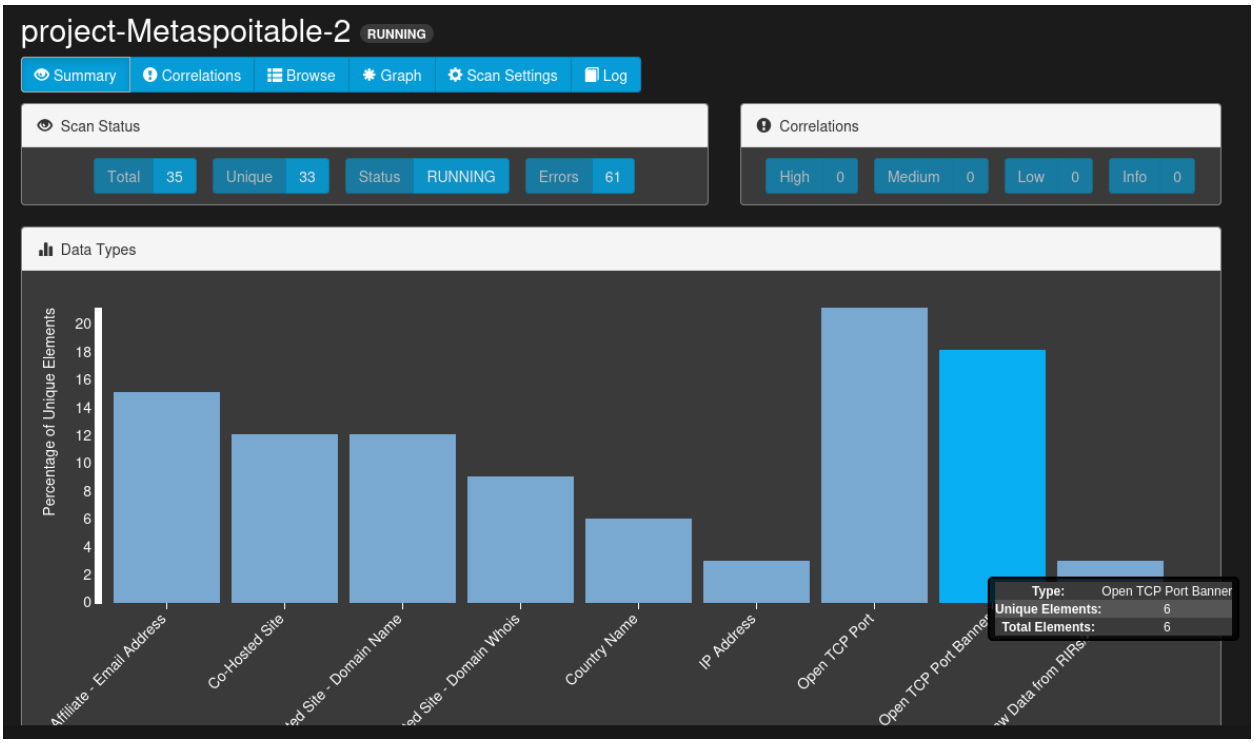
Real-time Monitoring

1. Navigate to **Scans** tab
2. Click on your active scan
3. View real-time results as they populate

Understanding the Results

Results are categorized into:

- **Entities**: Discovered assets (domains, IPs, emails, etc.)
- **Data Types**: Classification of findings
- **Modules**: Which module found the data
- **Correlation**: Relationships between entities



Analyzing Results

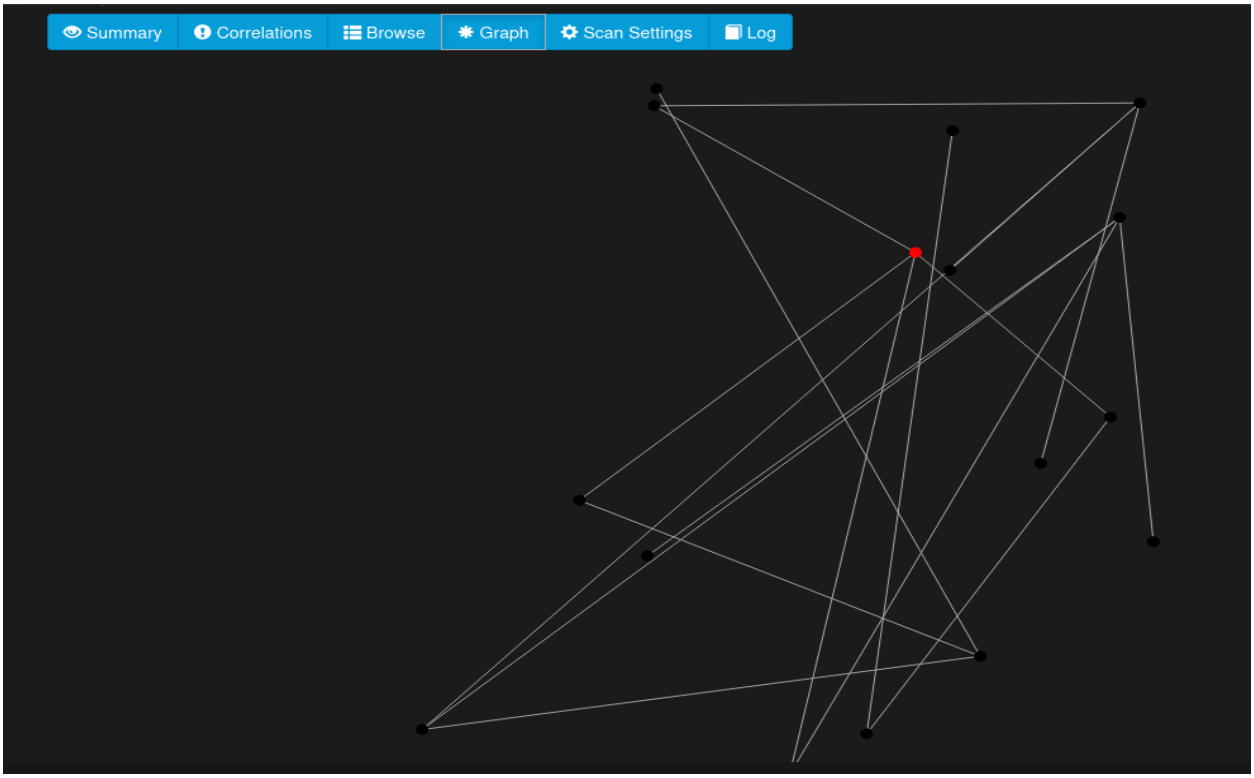
Step 1: Browse by Data Type

Click on different data categories:

- IP Addresses
- Domain Names
- Email Addresses
- Phone Numbers
- Physical Addresses
- Social Media Profiles
- Vulnerabilities

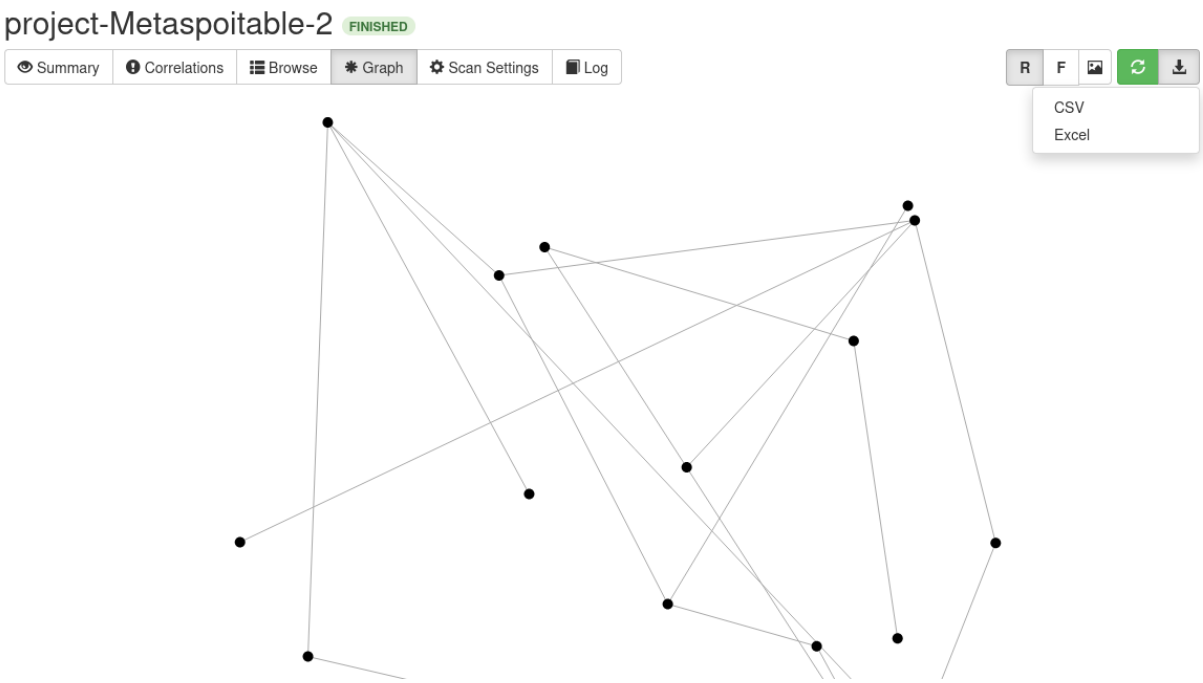
Step 2: View Relationships

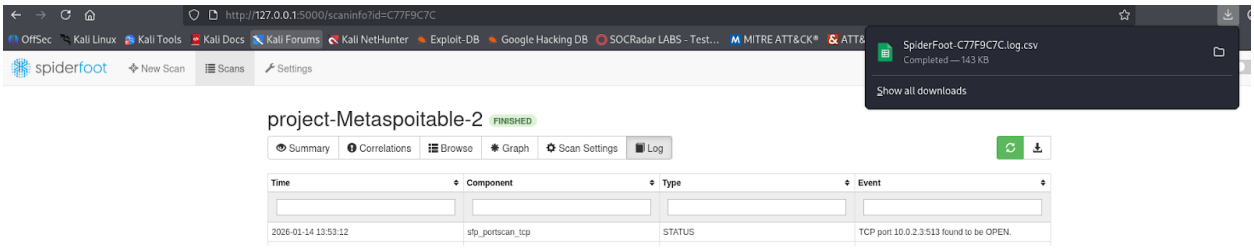
Click on **Graph** tab to see visual representation of relationships between discovered entities.



Step 3: Export Results

1. Click **Export** button
2. Choose format (JSON, CSV, GEXF)
3. Download file


project-Metaspoitable-2
FINISHED
Summary
Correlations
Browse
Graph
Scan Settings
Log
R
F
CSV
Excel


http://127.0.0.1:5000/scaninfo?id=C77F9C7C
OffSec
Kali Linux
Kali Tools
Kali Docs
Kali Forums
Kali NetHunter
Exploit-DB
Google Hacking DB
SOCRadar LABS - Test...
MITRE ATT&CK®
spiderfoot
New Scan
Scans
Settings
SpiderFoot-C77F9C7C.log.csv
Completed — 143 KB
Show all downloads
project-Metaspoitable-2
FINISHED
Summary
Correlations
Browse
Graph
Scan Settings
Log
Time
Component
Type
Event
2026-01-14 13:53:12
sfp_portscan_tcp
STATUS
TCP port 10.0.2.3:513 found to be OPEN.

Useful Command Options

Option	Description	Example
-s	Target to scan	-s example.com
-m	Modules to use	-m sfp_dnsresolve,sfp_whois
-o	Output format	-o json
-l	Listen address:port	-l 127.0.0.1:5001
-q	Quiet mode	-q
-t	Scan type	-t footprint

Practical Use Cases

Use Case 1: Company Security Assessment

Objective: Assess your organization's digital footprint

Steps:

1. Start a new scan with company domain
2. Select "All" scan type
3. Review discovered assets:
    - Subdomains
    - IP addresses
    - Email addresses
    - Exposed services
    - Security vulnerabilities

4. Identify shadow IT and unknown assets
5. Document findings for remediation.

Use Case 2: Threat Intelligence Gathering

Objective: Collect intelligence on potential threats

Steps:

1. Enter suspicious domain or IP
2. Enable threat intelligence modules
3. Check for:
    - Malware associations
    - Blacklist status
    - Historical DNS records
    - Related infrastructure
4. Export findings for threat analysis

Use Case 3: Pre-Engagement Reconnaissance

Objective: Gather information before penetration testing

Steps:

1. Create passive-only scan
2. Collect publicly available information
3. Map attack surface
4. Identify potential entry points
5. Document for penetration test planning

Use Case 4: Brand Monitoring

Objective: Monitor brand mentions and exposure

Steps:

1. Scan company name as target
2. Enable social media modules
3. Track:
    - Social media profiles
    - Mentions across platforms
    - Potential typosquatting domains
    - Unauthorized brand usage

Use Case 5: Data Breach Investigation

Objective: Determine if company data was exposed

Steps:

1. Enter company domain and email patterns
2. Enable breach data modules (HaveIBeenPwned, etc.)
3. Review compromised credentials
4. Cross-reference with employee database
5. Initiate password reset procedures.

Advantages and Disadvantages

Advantages

1. Comprehensive Data Collection

- Integrates 200+ data sources in one tool
- Eliminates need for multiple separate tools
- Provides holistic view of target

2. Automation and Efficiency

- Automates tedious manual OSINT tasks
- Saves hours or days of research time
- Runs multiple queries simultaneously

3. User-Friendly Interface

- Intuitive web-based GUI
- Visual data correlation and graphing
- Easy navigation for beginners

4. Customization and Flexibility

- Modular architecture allows selective scanning
- Customizable scan profiles
- CLI support for automation

5. Active Development and Community

- Regularly updated with new modules
- Active community support
- Open-source and free to use

6. Data Correlation Intelligence

- Automatically links related information
- Discovers hidden relationships
- Provides context to findings

7. Export and Reporting

- Multiple export formats
- Integration-ready outputs
- Shareable reports

8. No Direct Target Contact (Passive Mode)

- Can operate entirely passively
- Reduces detection risk
- Ethical reconnaissance option

Disadvantages

1. API Key Dependencies

- Many advanced features require API keys
- Some services have usage limits
- Cost associated with premium API access

2. Rate Limiting Issues

- Free API tiers have restrictions
- May require time between scans
- Can slow down comprehensive scans

3. Resource Intensive

- Full scans can take hours
- High memory usage with multiple modules
- Network bandwidth consumption

4. Learning Curve

- Understanding module purposes takes time
- Interpreting results requires OSINT knowledge
- Advanced configuration can be complex

5. False Positives

- Not all discovered data is relevant
- Requires manual verification
- Can generate information overload

6. Data Accuracy Concerns

- Relies on third-party data sources
- Information may be outdated
- No guarantee of data correctness

7. Legal and Ethical Risks

- Improper use can violate laws
- Requires understanding of legal boundaries
- Easy to accidentally cross ethical lines

8. Privacy Implications

- Can reveal sensitive information
- Potential for misuse
- Ethical considerations required

Best Practices

Security Best Practices

1. **Use in Controlled Environment**
    - Run on dedicated security testing systems
    - Isolate from production networks
    - Use VPN for additional privacy
2. **Secure Your Installation**
    - Use strong passwords for web interface
    - Restrict network access (use 127.0.0.1)
    - Keep SpiderFoot updated regularly
3. **Protect API Keys**
    - Store keys securely
    - Never commit keys to version control
    - Rotate keys periodically

Operational Best Practices

1. **Start with Passive Scans**
    - Begin with passive reconnaissance
    - Minimize target interaction
    - Reduce detection footprint
2. **Use Targeted Modules**
    - Don't always run "All" scans
    - Select relevant modules for your objective
    - Improves speed and reduces noise
3. **Document Your Findings**
    - Export results regularly
    - Maintain scan history

- Create reports for stakeholders

4. **Verify Critical Findings**
    - Manually verify important discoveries
    - Cross-reference with multiple sources
    - Don't rely solely on automated results
5. **Respect Rate Limits**
    - Space out comprehensive scans
    - Monitor API quota usage
    - Avoid overwhelming data sources

Ethical Best Practices

1. **Obtain Proper Authorization**
    - Only scan assets you own or have permission to test
    - Document authorization in writing
    - Respect scope boundaries
2. **Protect Discovered Data**
    - Handle sensitive findings responsibly
    - Encrypt exported data
    - Follow data protection regulations
3. **Be Transparent**
    - Disclose your activities to appropriate parties
    - Report vulnerabilities responsibly
    - Avoid deceptive practices

Legal and Ethical Considerations

Legal Framework

What is Legal

✅ **Permitted Activities:**

- Scanning your own assets and infrastructure
- Testing with written authorization
- Collecting publicly available information
- Security research on owned systems
- Educational purposes on test environments

What is Illegal

❌ **Prohibited Activities:**

- Unauthorized access to systems
- Scanning without permission
- Using discovered credentials without authorization
- Violating computer fraud and abuse laws
- Circumventing security controls

Key Legal Regulations

United States

- **Computer Fraud and Abuse Act (CFAA)**: Prohibits unauthorized access
- **Electronic Communications Privacy Act (ECPA)**: Protects electronic communications
- **State Laws**: Many states have additional cybersecurity laws

European Union

- **GDPR**: Governs personal data processing
- **Network and Information Security (NIS) Directive**: Security requirements
- **National Cybercrime Laws**: Vary by member state

International

- **Council of Europe Cybercrime Convention**: International framework
- **Local Laws**: Always research target country's regulations

Ethical Guidelines

Professional Ethics

1. **Informed Consent**: Always obtain explicit permission
2. **Scope Limitation**: Stay within authorized boundaries
3. **Data Protection**: Safeguard discovered information
4. **Responsible Disclosure**: Report vulnerabilities appropriately
5. **Transparency**: Be honest about your activities

Risk Assessment Questions

Before conducting any scan, ask yourself:

- Do I have explicit written permission?
- Am I authorized to access this information?
- Could this activity cause harm or disruption?
- Am I following organizational policies?
- Have I documented my authorization?

Responsible Disclosure

If you discover vulnerabilities:

1. **Document the Finding**
    - Record detailed information
    - Take screenshots as evidence
    - Note date and time
2. **Notify Affected Party**

- Contact the security team or the responsible person
- Provide clear, technical details
- Give reasonable time to remediate

3. **Don't Exploit**
    - Never use vulnerabilities for personal gain
    - Don't access data beyond what's necessary to confirm
    - Respect for embargo periods
4. **Follow Disclosure Policies**
    - Check if the organization has a bug bounty program
    - Follow coordinated disclosure timelines
    - Use encrypted communication

Conclusion

SpiderFoot is a powerful OSINT automation tool that significantly enhances the efficiency and comprehensiveness of intelligence gathering operations. When used responsibly and ethically, it provides invaluable insights for security assessments, threat intelligence, and reconnaissance activities.

Key Takeaways

- SpiderFoot automates OSINT collection from 200+ sources
- Offers both web GUI and CLI interfaces for flexibility
- Requires proper authorization and ethical use
- Provides comprehensive data correlation and visualization
- Active open-source community and regular updates